

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 26, 2003, 11:15:29 ; Search time 12235 Seconds  
(without alignments)  
30.445 Million cell updates/sec

Title: US-10-086-184-1

Perfect score: 23

Sequence: 1 aaatgcctccgagcgaggaac 23

Scoring table: IDENTITY\_NUC

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 60474

Minimum DB seq length: 0  
Maximum DB seq length: 40

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estrov:\*  
6: em\_estrpl:\*  
7: em\_estro:\*  
8: em\_hc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estcom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_trod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query	Match length	DB ID	Description
1	12.6	54.8	39	10	AM248192 2819671.5
2	12.4	53.9	34	9	AA938266 0097601.8
3	12.4	53.9	34	9	AI354718 4554112.x
4	12.2	53.0	28	9	AI646712 064407.x
5	12.2	53.0	37	9	AI966055 8c26d04.y
6	12.2	53.0	37	17	TA381B01Q

Result No.	Score	Query	Match length	DB ID	Description
7	12.2	53.0	40	9	AA939192 0q22h09.8
8	11.8	51.3	37	9	AI474098 1A174098 1m05c07.x
9	11.8	51.3	39	17	A2430589 1M0215F10
10	11.6	50.4	38	17	TA137C040
11	11.6	50.4	39	17	AZ800550
12	11.4	49.6	36	9	AA825458
13	11.4	49.6	37	14	H43693
14	11.2	48.7	27	9	AU254523
15	11.2	48.7	27	14	D18732
16	11.2	48.7	31	9	AA920864
17	11.2	48.7	37	13	B1561770
18	11.2	48.7	38	13	B1561718
19	10.8	47.0	22	9	AA985475
20	10.8	47.0	29	14	H55186
21	10.8	47.0	32	17	A2813220
22	10.8	47.0	33	17	A2803201
23	10.8	47.0	34	9	AA517484
24	10.8	47.0	35	2	HSN009944
25	10.8	47.0	35	17	A2469734
26	10.8	47.0	35	17	A2599526
27	10.8	47.0	36	12	BF533462
28	10.8	47.0	37	9	AI191467
29	10.8	47.0	37	9	AI735401
30	10.8	47.0	37	17	TA168H1P
31	10.6	46.1	34	10	BE386585
32	10.6	46.1	35	13	B1827700
33	10.6	46.1	37	9	AI538439
34	10.6	46.1	38	12	BF687876
35	10.6	46.1	39	17	BH866488
36	10.6	46.1	40	9	AI697005
37	10.4	45.2	26	17	BH759478
38	10.4	45.2	34	9	AA989545
39	10.4	45.2	34	10	AV833436
40	10.4	45.2	37	10	AV958431
41	10.4	45.2	37	17	TA21B11P
42	10.4	45.2	40	9	AI424339
43	10.4	45.2	40	17	BH857109
44	10.2	44.3	29	10	AW250348
45	10.2	44.3	32	12	BG718358

## ALIGNMENTS

RESULT 1  
LOCUS AM248192 39 bp mRNA linear EST 07-JUN-2000  
DEFINITION 2819671.5prime NIH\_MGC\_7 Homo sapiens cDNA clone IMAGE:2819671 5',  
ACCESSION AM248192  
VERSION AM248192.1 GI:6591185  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 39)  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.  
TITLES Unpublished (1999)  
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
COMMENT Other ESTs: 2819671.3prime  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: DCTD/DP CDNA Library Preparation: Ling  
Hong/Rubin Laboratory CDNA Library Arayed by: The I.M.A.G.E.  
Consortium (LNL) DNA Sequencing by: Berkeley MGC Sequencing  
Project Clone distribution: MGC clone distribution information can  
be found through the I.M.A.G.E. Consortium/LNL at:  
www.bio.lnlnl.gov/bbrp/image/image.html Base Calling / Quality  
Scores: PHRED from University of Washington Genome Center. Vector  
Trimming: cross\_match from University of Washington Genome Center  
PHRAP suite. Poly-T Identification: patmatch.pl from Berkeley  
Drosophila Genome Project. University of Washington Genome Center:

http://www.genome.washington.edu/LowQualitySequence: 0 contiguous  
PHRED high quality bases following vector sequence. Very low  
quality sequence: Trace file contained 39 contiguous distinct peaks  
following vector sequence.  
Plate: L1CM2 row: D column: 8.

## FEATURES

source

1.39  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="IMAGE:2819671"  
/clone\_1ib="NIH MGC 7"  
/cissue\_type="small cell carcinoma"  
/cell\_line="MGC3"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: Lung; Vector: pOTB1; Site 1: XhoI; Site 2:  
EcoRI; cDNA made by oligo-dT priming. Directionally  
cloned into EcoRI/XhoI sites using the following 5'  
adapter: GGACGAG(G). Size-selected >500bp for average  
insert size 1.8kb. Library constructed by Ling Hong in  
the Laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies)."

## BASE COUNT

7 a 11 c 17 g 4 t

ORIGIN

Query Match 54.8%; Score 12.6; DB 10; Length 39;  
Best Local Similarity 78.9%; Pred. No. 1.5e+05;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 AAATCGGCTCGAGCGGAG 19

Db 14 AAGCGCGCTGAGCGCGG 32

RESULT 2  
AA938266 25 bp mRNA linear EST 27-AUG-1998  
LOCUS 0097601.81 NCI CGAP KIDS Homo sapiens cDNA clone IMAGE:1574136 3'  
DEFINITION similar to SW:CA12\_MOUSE P28481 PROCOLLAGEN ALPHA 1(I) CHAIN  
PRECEDSOR : mRNA sequence.

ACCESSION AA938266  
VERSION AA938266  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 25)  
NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov

## JOURNAL

COMMENT

tisue Procurement: Christopher Moskalkuk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CCGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bdrip/image/image.html

Trace considered overall poor quality  
Insert Length: 283 Std Error: 0.00  
Seq primer: -40ml3 fwd. RT from Amersham  
High quality sequence stop: 1.  
Location/Qualifiers

## FEATURES

source

1.25  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1574136"  
/clone\_1ib="NCI\_CGAP\_KIDS"

## BASE COUNT

7 a 8 c 10 g 0 t

ORIGIN

Query Match 53.9%; Score 12.4; DB 9; Length 25;  
Best Local Similarity 72.7%; Pred. No. 1.7e+05;  
Matches 16; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 1 AAATCGGCTCGAGCGGAGAA 22

Db 4 AAGCGCGCGCGCGGAGAA 25

RESULT 3  
A1354718 34 bp mRNA linear EST 04-JAN-1999  
LOCUS qt54h12.x1 Soares fetal\_lung NBH119W Homo sapiens cDNA clone  
DEFINITION IMAGE:1951847 3' similar to SW:PRPB\_HUMAN P2814 PROLINE-RICH  
PEPTIDE P-B.; contains element WSH1 repetitive element ;, mRNA  
sequence.

ACCESSION A1354718  
VERSION A1354718  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 34)  
NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov

## JOURNAL

COMMENT

This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Trace considered overall poor quality  
Seq primer: -40UP from Gibco  
High quality sequence stop: 1.  
Location/Qualifiers

## FEATURES

source

1.34  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1951847"  
/clone\_1ib="Soares fetal\_lung\_NBH119W"  
/dev stage="19 weeks"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: Lung; Vector: pT73D (Pharmacia) with a  
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer  
15'-GTGTACCATCTGAGTGGAGCGCGCCCAATTTTCTTTT-3',  
double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified pT73 vector  
(Pharmacia). Library went through one round of  
normalization to a Cot = 5. Library constructed by Bento  
Soares and M. Fatima Bonaldo. This library was constructed  
from the same fetus as the fetal heart library, Soares  
fetal heart NBH119W."

## BASE COUNT

8 a 7 c 15 g 4 t

ORIGIN

Query Match 53.9%; Score 12.4; DB 9; Length 34;

Best Local Similarity 72.7%; Pred. No. 1.7e+05;  
Matches 16; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 AAATCGGCTCCGAGCGGGA 22  
1 AAAAAAGGTTCCGCGGGGGA 22  
Db 1 AAAAAAGGTTCCGCGGGGGA 22

RESULT 4  
A1646712 28 bp mRNA linear EST 29-APR-1999  
LOCUS  
DEFINITION  
IMAGE:138557 3' similar to TR:Q08805 Q08805 SALIVARY PROLINE-RICH  
PROTEIN L.; mRNA sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
EST.  
house mouse.  
Mus musculus

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.  
Unpublished (1997)  
Tumor Gene Index  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
This clone is available royally-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:905025  
This clone was previously sequenced on the 5' end only, this new  
data is from the 3' end  
Trace considered overall poor quality  
Possible reversed clone; similarity on wrong strand  
High quality sequence stop: 1.  
Location/Qualifiers

FEATURES  
source  
1..28  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="IMAGE:138257"  
/clone\_11b="Soares\_mammary\_gland\_NMLMG"  
/sex="female (lactating)"  
/tissue\_type="mammary gland"  
/lab\_host="DH10B"  
/note="vector: pT7T3D-Pac (Pharmacia) with a modified  
polylinker; 1st strand cDNA was prepared from mammary  
gland tissue from a lactating female, and was then primed  
with a Not I - oligo(dT) primer. Double-stranded cDNA was  
ligated to Eco RI adaptors (Pharmacia), digested with Not  
I and cloned into the Not I and Eco RI sites of the  
modified pT7T3 vector. Library is normalized. Library  
was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 10 a 9 c 8 g 1 t  
ORIGIN

Query Match 53.0%; Score 12.2; DB 9; Length 28;  
Best Local Similarity 82.4%; Pred. No. 2.1e+05;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 6 GGCTCCGAGCGGGA 22  
5 GGCCCCAGGGGGA 21  
Db 5 GGCCCCAGGGGGA 21

RESULT 5  
A1966055/c 37 bp mRNA linear EST 30-NOV-2001  
LOCUS  
DEFINITION  
IMAGE:138557 3' similar to TR:Q08805 Q08805 SALIVARY PROLINE-RICH  
PROTEIN L.; mRNA sequence.

KEYWORDS  
SOURCE  
ORGANISM  
EST.  
soybean.  
Glycine max

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots;  
Rosidae; eucosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;  
Glycine.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
Public Soybean EST Project  
Unpublished (1999)  
Contact: Shoemaker R/Public Soybean EST Project  
Public Soybean EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu

Trace considered overall poor quality  
Possible reversed clone; similarity on wrong strand This clone is  
available through: Resgen, Invitrogen Corp. 2130 South Memorial  
Parkway Huntville, AL 35801 For further information call: (800  
)-533-4363 or contact via email: cse@resgen.com  
Seq primer: -40RP from Gibco  
High quality sequence stop: 1.  
Location/Qualifiers

FEATURES  
source  
1..37  
/organism="Glycine max"  
/db\_xref="taxon:3847"  
/clone="GENOME SYSTEMS CLONE ID: Gm-c1013-1760"  
/clone\_11b="Gm-c1013"  
/tissue\_type="whole seedlings, 2-3 week old seedlings,  
greenhouse grown"  
/lab\_host="XL10-Gold"  
/note="vector: pBluescript II XR; Site 1: EcoRI; Site 2:  
XhoI; This cDNA library was constructed from mRNA isolated  
from whole seedlings of 2-3 week old greenhouse grown  
plants. The cDNA library was prepared using the Stratagene  
pBluescript II XR cDNA library construction kit.  
Complementary DNA was synthesized from mRNA using a primer  
consisting of a poly (dT) sequence with a XhoI restriction  
site. EcoRI adaptors were ligated to the blunt-ended cDNA  
fragments followed by XhoI digestion. The cDNA fragments  
were directionally cloned into the EcoRI-XhoI restriction  
site of the pBluescript vector. The ligated cDNA fragments  
were transformed into XL10-Gold host cells. This library  
was constructed by Dr. Randy Shoemaker and Dr. John  
Erpelding."

BASE COUNT 7 a 18 c 2 g 10 t  
ORIGIN

Query Match 53.0%; Score 12.2; DB 9; Length 37;  
Best Local Similarity 82.4%; Pred. No. 2.2e+05;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 6 GGCTCCGAGCGGGA 22  
36 GGCTTCGAGGTGGA 20  
Db 36 GGCTTCGAGGTGGA 20

RESULT 6  
TA381B010/c 37 bp DNA linear GSS 13-DEC-2000  
LOCUS  
DEFINITION  
T. brucei sheared genomic DNA clone 381B01, reverse sequence.  
genomic survey sequence.

SOURCE Trypanosoma brucei.  
ORGANISM Trypanosoma brucei  
REFERENCE Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
AUTHORS 1 (bases 1 to 37)  
Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R., Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L., Melville, S.E., Rajandream, M.A. and Barrell, B.G.  
TITLE Direct Substitution  
JOURNAL Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and nhleanger@ac.uk  
COMMENT Constructed at the Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared to give a tight size distribution (4 kb). The v + i method used for the library construction is described in detail in Smith, H. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaubin and B. Barrell, Oxford University Press, 1999).  
Email: nhleaser@tigr.org  
Details of T. brucei sequencing at the Sanger Centre are available at [http://www.sanger.ac.uk/Projects/T\\_brucei/](http://www.sanger.ac.uk/Projects/T_brucei/).  
Location/Qualifiers  
1. 37  
/organism="Trypanosoma brucei"  
/strain="TREU927"  
/db\_xref="taxon:5691"  
/clone="381b01"  
BASE COUNT 5 a 22 c 9 g 1 t  
ORIGIN  
Query Match 53.0%; Score 12.2; DB 17; Length 37;  
Best Local Similarity 82.4%; Pred. No. 2.2e+05;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 CGGCTCCGAGCGGGAA 21  
|||||  
Db 25 CGGCGCCGCGCGGCA 9

RESULT 7  
AA939192/c 40 bp mRNA linear EST 01-MAY-1998  
LOCUS oq2209.g1 NCI CGAP GC4 Homo sapiens cDNA clone IMAGE:1587137.3  
DEFINITION similar to SW-PRPE HUMAN P02811 BASIC PROLINE-RICH PEPTIDE P-E  
AA939192  
/contains KER.b2 MER30 repetitive element ;, mRNA sequence.  
ACCESSION AA939192.1 GI:3099105  
VERSION  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 40)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
COMMENT Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: [www-bio.llnl.gov/bdrp/image/image.html](http://www-bio.llnl.gov/bdrp/image/image.html)  
Trace considered overall poor quality

Seq primer: -40m13 fwd. Et from Amersham  
High quality sequence scop: 1.  
Location/Qualifiers  
1. 40  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1587137"  
/clone\_lib="NCI CGAP GC4"  
/tissue\_type="pooled\_germ cell tumors"  
/lab\_host="DH108"  
/note="Vector: pTR73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from 3 pooled germ cell tumors, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pTR73 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."  
BASE COUNT 6 a 23 c 6 g 5 t  
ORIGIN  
Query Match 53.0%; Score 12.2; DB 9; Length 40;  
Best Local Similarity 82.4%; Pred. No. 2.2e+05;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 6 GGCTCCGAGCGGGAA 22  
|||||  
Db 32 GGCCCGCGGGGAA 16

RESULT 8  
A1474098 37 bp mRNA linear EST 14-APR-1999  
LOCUS tm05c07.x1 NCI CGAP Col4 Homo sapiens cDNA clone IMAGE:215562.3  
DEFINITION similar to TR:Q39865 Q39865 HYDROXYPROLINE-RICH GLYCOPROTEIN  
A1474098  
/contains element MSRI repetitive element ;, mRNA sequence.  
ACCESSION A1474098.1 GI:4327143  
VERSION  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 37)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
COMMENT Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: [www-bio.llnl.gov/bdrp/image/image.html](http://www-bio.llnl.gov/bdrp/image/image.html)  
Trace considered overall poor quality  
Insert Length: 968 Std Error: 0.00  
Seq primer: -40UP from Gibco  
High quality sequence scop: 1.  
Location/Qualifiers  
1. 37  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:215562"  
/clone\_lib="NCI CGAP Col4"  
/tissue\_type="moderately-differentiated adenocarcinoma"  
/lab\_host="DH108"  
/note="Organ: colon; Vector: PCMW-SPORT6; Site 1: SalI; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT."  
Trace considered overall poor quality

Average insert size 1.7 kb. Life Technologies catalog #:  
11531-019"

Query Match 51.3%; Score 11.8; DB 9; Length 37;  
Best Local Similarity 69.6%; Pred. No. 3.2e+05;  
Matches 16; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 AAATGGCTCCGAGCGGGAAC 23  
Db 1 AAGGGGCCCCGGGGGCGCAAC 23

RESULT 9  
LOCUS AZ430589 39 bp DNA linear GSS 03-OCT-2000  
DEFINITION M0215F10F Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
clone UUGC1M0215F10 F, DNA sequence.

ACCESSION AZ430589  
VERSION AZ430589.1 GI:10554602  
KEYWORDS GSS

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 39)

AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Haml, C.,  
Islam, H., Longacre, S., Mahmud, M., Meenen, E., Pedersen, T., Reilly,  
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A.  
and Wright, D., Weiss, R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts

JOURNAL Unpublished (2000)

COMMENT Contact: Robert B. Weiss  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA

Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0215 row: F column: 10  
Seq primer: CGTGTGAACGACGCGCAGT  
Class: plasmid ends  
High quality sequence stop: 39.

FEATURES  
Location/Qualifiers

1..39

/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC1M0215F10"  
/clone\_11b="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: pMD42ny. Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adapted DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of pMD42 (gi|4732114|gb|AF129072.1), a copy-number  
inducible derivative of plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adapted mouse DNA was annealed to  
adapted vector DNA, and transformed into

chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."

Query Match 51.3%; Score 11.8; DB 17; Length 39;  
Best Local Similarity 86.7%; Pred. No. 3.2e+05;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 5 CGGCTCCGAGCGGCGG 19  
Db 13 CGCTCCAGCGCGG 27

RESULT 10  
LOCUS TA137C04Q 38 bp DNA linear GSS 13-DEC-2000  
DEFINITION T. brucei sheared genomic DNA clone 137C04, reverse sequence,  
genomic survey sequence.

ACCESSION AL466262  
VERSION AL466262.1 GI:11835617  
KEYWORDS GSS

SOURCE Trypanosoma brucei.

ORGANISM Trypanosoma brucei

REFERENCE Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;  
Trypanosoma.  
1 (bases 1 to 38)

AUTHORS Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R.,  
Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L.,  
Melville, S.E., Rajandream, M.A. and Barrell, B.G.

TITLE Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing  
project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,  
Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and  
nh@sanger.ac.uk

COMMENT Constructed at the Institute for Genomic Research (TIGR),  
Rockville, MD. Genomic DNA isolated from a cloned population of  
Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared  
to give a tight size distribution (4 kb). The v + 1 method used for the library construction is  
described in detail in Smith, H. and Venter, J.C. (Making small  
insert libraries for whole genome shotgun sequencing projects. In  
Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.  
Barrell, Oxford University Press, 1999).

Email: nh@sanger.ac.uk  
Details of T. brucei sequencing at the Sanger Centre are available  
at http://www.sanger.ac.uk/Projects/T\_brucei/.

FEATURES  
Location/Qualifiers

1..38

/organism="Trypanosoma brucei"  
/strain="TREU927"  
/db\_xref="taxon:5691"  
/clone="137C04"  
BASE COUNT 9 a 11 c 14 g 4 t

Query Match 50.4%; Score 11.6; DB 17; Length 38;  
Best Local Similarity 77.8%; Pred. No. 3.9e+05;  
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 5 CGGCTCCGAGCGGGA 22  
Db 18 CAGCAACGAGCGCGG 35

RESULT 11  
LOCUS AZ800550 39 bp DNA linear GSS 16-FEB-2001  
DEFINITION 2M0058810R Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
clone UUGC2M0058810 R, DNA sequence.

ACCESSION AZ800550  
VERSION AZ800550.1 GI:12952808  
KEYWORDS GSS.

SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 39)  
AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamli, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A., and Wright, D., Weiss, R.  
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
JOURNAL Unpublished (2000)  
COMMENT Contact: Robert B. Weiss  
University of Utah Genome Center  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0058 row: E column: 10  
Seq primer: CACACAGGAACGCTATGACC  
Class: plasmid ends  
High quality sequence stop: 39.  
Location/Qualifiers  
1. 39  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="U06C2M0058E10"  
/clone\_1b="Mouse 10kb plasmid U06C2M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: pMD22ny. Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD22 (g14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 3 a 12 c 18 g 6 t  
ORIGIN  
Query Match 50.4%; Score 11.6; DB 17; Length 39;  
Best Local Similarity 77.8%; Pred. No. 3.9e+05;  
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 6 GGCTCCGAGCGCGGAAC 23  
DB 28 GGCCCCCGAGCGGAGACC 11

RESULT 12  
LOCUS AA825458 36 bp mRNA linear EST 24-FEB-1998  
DEFINITION o664f12.81 NCI CGAP L45 Homo sapiens cDNA clone IMAGE:1416431 3'  
ACCESSION AA825458  
VERSION AA825458.1 GI:2898770  
KEYWORDS EST.

SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 36)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapb-r@mail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bcrp/image/image.html

Trace considered overall poor quality  
Seq primer: -40m13 fwd. RT from Amersham  
High quality sequence stop: 1.  
Location/Qualifiers  
1. 36  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1416431"  
/clone\_1b="NCI CGAP L45"  
/tissue\_type="Carcinoid"  
/lab\_host="DH10B"  
/note="Organ: lung; Vector: p773D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from neuroendocrine lung carcinoid, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p773 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 4 a 8 c 9 g 15 t  
ORIGIN  
Query Match 49.6%; Score 11.4; DB 9; Length 36;  
Best Local Similarity 71.4%; Pred. No. 4.7e+05;  
Matches 15; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 3 ATCGGCTCCGAGCGCGGAAC 23  
DB 8 ATCTGCTTCGTGCGGTTAC 28

RESULT 13  
LOCUS H43693 37 bp mRNA linear EST 31-JUL-1995  
DEFINITION y080b05.s1 Soares adult brain N2b4H85Y Homo sapiens cDNA clone IMAGE:184209 3' similar to SP:R55\_RAT P24050 40S RIBOSOMAL PROTEIN 1, mRNA sequence.  
ACCESSION H43693  
VERSION H43693.1 GI:919745  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 37)  
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P., and Wilson, R.  
TITLE The WashU-Merck EST Project  
JOURNAL Unpublished (1995)  
COMMENT Contact: Wilson RK  
Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu

Insert Size: 851  
High quality sequence starts: 1  
High quality sequence stops: 1  
Source: IMAGE Consortium, LNL  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
Trace considered overall poor quality  
Possible reversed clone: similarity on wrong strand  
Insert Length: 851 Std Error: 0.00  
Seq primer: Promega -21m13  
High quality sequence stop: 1.  
Location/Qualifiers

# FEATURES

source

```
1..37
/organism="Homo sapiens"
/db_xref="DB:3828558"
/db_xref="taxon:9606"
/clone="IMAGE:184209"
/clone_lib="Soares adult brain N2bH855Y"
/sex="Male"
/dev_stage="55-year old"
/lab_host="DH10B (ampicillin resistant)"
/notes="Organ: brain; Vector: pRT73D (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer (5'
TGTTCACATCTGAAGTGGAGCGCGCGCTTTTCTTTTCTTTT 3').
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pRT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 53. Library constructed by Bento
Soares and M.Fatima Bonaldo. The adult brain RNA was
provided by Dr. Donald H. Gilden. Tissue was acquired
17-18 hours after death which occurred in consequence of a
ruptured aortic aneurysm. RNA was prepared from a pool of
tissues representing the following areas of the brain:
frontal, parietal, temporal and occipital cortex from the
left and right hemispheres, subcortical white matter,
basal ganglia, thalamus, cerebellum, midbrain, pons and
medulla."
```

BASE COUNT 6 a 8 c 11 g 10 t 2 others  
ORIGIN

Query Match 49.6%; Score 11.4; DB 14; Length 37;  
Best Local Similarity 65.2%; Pred. No. 4.8e+05;  
Matches 15; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 AATCGGCTCCGAGCGGGAAC 23  
DB 32 ACACCTGCTNANNGCGAGAAC 10

RESULT 14  
AUS254523 27 bp mRNA linear EST 25-APR-2002  
LOCUS AUS254523 3'-directed mouse cDNA library Mus musculus cDNA clone  
DEFINITION BED0002375 3', mRNA sequence.  
ACCESSION AU254523  
VERSION AU254523.1 GI:20316384  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 27)  
Kato, K. and Matoba, R.  
Generation of expressed sequence tags from mouse brain  
TITLE Unpublished (2002)  
JOURNAL  
COMMENT Graduate School of Biological Sciences

Nara Institute of Science and Technology  
8916-5 Takayama, Ikoma, Nara 630-0101, Japan  
Tel: 81-743-72-5581  
Fax: 81-743-72-5589  
Email: kkat@nara.ac.jp, BED/index.html.  
URL: http://love2.aist-nara.ac.jp/BED/index.html.  
Location/Qualifiers

# FEATURES

source

```
1..27
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="BED0002375"
/clone_lib="3'-directed mouse cDNA library"
/tissue_type="brain"
/notes="Vector: pCEM-T-easy"
BASE COUNT 7 a 5 c 9 g 6 t  
ORIGIN
```

Query Match 48.7%; Score 11.2; DB 9; Length 27;  
Best Local Similarity 81.2%; Pred. No. 5.6e+05;  
Matches 13; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 7 GCTCCGAGCGGGAAC 22  
DB 12 GCTCTGAGCAGAGAA 27

RESULT 15  
D18732 27 bp mRNA linear EST 12-DEC-1995  
LOCUS D18732  
DEFINITION MUS5501794 Mouse 3'-directed Mus musculus domesticus cDNA clone  
md0354 3', mRNA sequence.  
ACCESSION D18732  
VERSION D18732.1 GI:1100701  
KEYWORDS EST.  
SOURCE western European house mouse.  
ORGANISM Mus musculus domesticus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 27)  
Kawamoto, S., Okubo, K., Yoshii, J., Katsuki, M. and Matsubara, K.  
Analysis of gene expression in mouse embryogenesis by 3'-directed  
cDNA sequencing  
TITLE Unpublished (1995)  
JOURNAL  
COMMENT Contact: Kawamoto, S., Okubo, K., Yoshii, J., Katsuki, M. and Matsubara  
K.  
Institute for Cellular and Molecular Biology  
Osaka University  
3-1 Yamada-oka, Suita, Osaka 565, Japan.  
Location/Qualifiers

# FEATURES

source

```
1..27
/organism="Mus musculus domesticus"
/strain="C57BL/6J"
/db_xref="taxon:10092"
/clone="md0354"
/clone_lib="Mouse 3'-directed"
/tissue_type="decidual tissue (day 6.5-8.5 of gestation)"
BASE COUNT 7 a 5 c 9 g 6 t  
ORIGIN
```

Query Match 48.7%; Score 11.2; DB 14; Length 27;  
Best Local Similarity 81.2%; Pred. No. 5.6e+05;  
Matches 13; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 7 GCTCCGAGCGGGAAC 22  
DB 12 GCTCTGAGCAGAGAA 27

Search completed: March 26, 2003, 23:17:58  
Job time : 12237 secs

**THIS PAGE BLANK (USPTO)**